Cramér-Rao Bounds for Polynomial Signal Estimation using Sensors with AR(1) Drift

Swarnendu Kar, *Student Member, IEEE,* Pramod K. Varshney, *Fellow, IEEE,*
and Marimuthu Palaniswami, *Fellow, IEEE*

Abstract—We seek to characterize the estimation performance of a sensor network where the individual sensors exhibit the phenomenon of drift, i.e., a gradual change of the bias. Though estimation in the presence of random errors has been extensively studied in the literature, the loss of estimation performance due to systematic errors like drift have rarely been looked into. In this paper, we derive closed-form Fisher Information matrix and subsequently Cramér-Rao bounds (upto reasonable approximation) for the estimation accuracy of drift-corrupted signals. We assume a polynomial time-series as the representative signal and an autoregressive process model for the drift. When the Markov parameter for drift $\rho < 1$, we show that the first-order effect of drift is asymptotically equivalent to scaling the measurement noise by an appropriate factor. For $\rho = 1$, i.e., when the drift is non-stationary, we show that the constant part of a signal can only be estimated inconsistently (non-zero asymptotic variance). Practical usage of the results are demonstrated through the analysis of 1) networks with multiple sensors and 2) bandwidth limited networks communicating only quantized observations.

Index Terms—Sensor Networks, Systematic Errors, Autoregressive Process, Polynomial Regression, Distributed Estimation.

I. INTRODUCTION

A. The phenomenon of drift

Sensor networks often consist of a number of sensors deployed for specific tasks such as detection and estimation of the ambient phenomenon [1]. Sensing nodes are equipped with sensing, computation, and inference capabilities. Sensor networks can potentially deploy a large number of sensors, can measure the environment over time and generate data, from which we seek to extract relevant information regarding the phenomenon. Often while assessing the estimation performance when using such a network, sensor inaccuracies are modeled as independent measurement errors, e.g., i.i.d. Gaussian errors were assumed in [2],[3]. However, it has been widely reported that a variety of sensors used in various applications exhibit systematic errors or drift, e.g., Ground Moving Target Indicator radar sensors [4], tilt sensors in bridge

S. Kar and P. K. Varshney are with the Department of Electrical Engineering and Computer Science, Syracuse University, NY 13244 USA, e-mail: {swkar,varshney}@syr.edu

M. Palaniswami is with the Department of Electrical and Electronic Engineering, The University of Melbourne, Victoria 3010 Australia, email: palani@unimelb.edu.au

monitoring applications [5], CO_2 sensors used for air quality monitoring [6] and salinity sensors for ocean monitoring [7]. For example, gas sensors are known to drift due to other environmental parameters like temperature and humidity [8].

B. Inference with drifting sensors

When sensors in a network develop systematic errors with time, inferences based on these observations become increasingly inaccurate. For example, a scheme where each sensor can track their own drift in a collaborative manner is presented in [9] and such a system was shown to accumulate error over time. This necessitates periodic calibration/ registration of the sensors, a procedure that increases the operating cost for the network. Though the issue of drift is often acknowledged in the literature, performance analysis of a network consisting of drifting sensors has not been done before. Our goal in this paper is to characterize the estimation performance of a sensor network in terms of the drift properties of the constituent sensors.

We consider multiple sensors deployed in a field of interest to monitor a particular environment. The sensors observe the same phenomena over $n = 1, 2, \ldots, N$ time instants and their observations are corrupted by noise and drift. The noisy observations are relayed to a common sink, where the objective is to estimate the parameters governing the observed phenomena. Let $z_{n,m}$ denote the noisy observation of the mth sensor at the nth instant and $\boldsymbol{z}_m$ denote the vector $[z_{1,m}, z_{2,m}, \ldots, z_{N,m}]'$. We assume a linear observation model of the form

$$\boldsymbol{z}_m = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}_m, \quad \boldsymbol{\epsilon}_m \sim \mathcal{N}(0, \boldsymbol{\Sigma}_m) \tag{1}$$

where $\boldsymbol{x} \triangleq \boldsymbol{X}\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}_m$ are $N \times 1$ vectors denoting the signal and error terms. Here $\boldsymbol{X}$ is assumed to be a $N \times (P+1)$ known matrix (describing the temporal-dynamics) and $\boldsymbol{\beta}$ is the $(P + 1) \times 1$ vector denoting an unknown but deterministic signal. The error term $\boldsymbol{\epsilon}_m$ consists of noise and drift components (to be described later) and is assumed to be Gaussian distributed with covariance matrix $\boldsymbol{\Sigma}_m$. The observations $\boldsymbol{z}_m$ are communicated to a sink, whose job is to obtain an accurate estimate of β. From estimation theory (see, for example, [10] and [11]), it is well known that the variance of an unbiased estimator is lower bounded by the Cramér-Rao Lower Bound (CRB) and that the Maximum-Likelihood (ML) estimator asymptotically (for small errors) attains that bound. In this paper, we derive the closed-form CRB (upto reasonable approximations) for estimating $\boldsymbol{\beta}$ when $\boldsymbol{\Sigma}_m$ corresponds to drift corrupted errors and $\boldsymbol{X}$ corresponds to a polynomial

signal. In the subsequent discussion, we provide the motivation and full descriptions of $\boldsymbol{\Sigma}_m$ and $\boldsymbol{X}$.

C. Models of Noise and Drift

At instant n, let x_n and $\epsilon_{n,m}$ denote the common signal magnitude and observation error of sensor m. When the effect due to drift is ignored, the measurement noise is often modeled as independent and identically distributed Gaussian noise $w_{n,m}$, (e.g., [2], [3]), i.e.,

$$z_{n,m} = x_n + w_{n,m}, \quad w_{n,m} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_m^2), \qquad (2)$$

where σ_m^2 is the noise variance. However, in the presence of drift, the observation error has two components: one due to baseline-drift $d_{n,m}$ and the other due to random measurement noise $w_{n,m}$,

$$z_{n,m} = x_n + d_{n,m} + w_{n,m}. \qquad (3)$$

Drift is generally described as a gradual change of the bias of a sensor [8], [12]. Depending on the specific sensing methodology, various models have been proposed to characterize the drift sequence $\{d_{n,m}\}_{n=1}^N$ in frequency and time domains (for a survey, see [8]). Below we describe three commonly used models of drift:

1) Frequency domain: The phase-drift in an oscillator is often modeled in the frequency domain using several powers of frequency (power-law model) [12], i.e., the power spectral density (PSD) is of the form, $S_m(f) = \sum_{i=-2}^{2} h_{m,i} f^i$, where $h_{m,i}$ are appropriate constants.

2) Temporal domain - Deterministic: Drift is also modeled sometimes as a linear movement of the sensor baseline, i.e., $d_{n,m} = a_m + (n-1)b_m$, where the intercept a_m is set to zero after each calibration and slope b_m is assumed to be an unknown constant that is often estimated later and compensated for. Example applications include odor identification using gas-sensor arrays [13] and air pollution monitoring using gas sensor networks [14].

3) Temporal domain - Random: In the broader signal processing literature, auto-regressive moving average (ARMA) processes are often used to describe serially correlated time series, an example of which is drift. As a trade-off between modeling-efficiency and analytical-complexity, drift is often modeled as a first-order autoregressive (AR(1)) process. Example applications include Ground Moving Target Indicator (GMTI) radars ([4], [15]), Ring-laser Gyroscopes [16], Liquid Chromatography [17] and sensor networks [9].

In this paper, we will use the first-order autoregressive model to describe the statistical properties of drift. The AR(1) model characterizes the drift behavior at m^{th} sensor in terms of an auto-correlation parameter ρ_m (visually, a smaller value of ρ_m means that the baseline drift crosses the zero-line more often and looks more like white noise) and a strength parameter $\sigma_{\delta,m}^2$ (signifying the magnitude of drift),

$$d_{n+1,m} = \rho_m d_{n,m} + \delta_{n,m}, \text{ where}$$
$$\rho_m \in [0,1] \text{ and } \delta_{n,m} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_{\delta,m}^2). \qquad (4)$$

Note that, when $\rho = 1$, the drift is modeled effectively as a non-stationary random walk, as in [9].

If we define γ_m such that $\sigma_{\delta,m}^2 = \gamma_m \sigma_m^2$, our "AR(1)+White noise" model for observation error is completely parameterized by $\{\sigma_m^2, \rho_m, \gamma_m\}$. These noise and drift parameters usually have to be identified from the empirical PSD for stationary noise (e.g., [17], [18]) or other time-domain features for non-stationary noise (e.g., [19]). In this paper, we consider the characterization of the sensing uncertainty in terms of $\{\sigma_m^2, \rho_m, \gamma_m\}$ as part of system identification that must be done prior to an observation cycle. Estimation of these parameters is beyond the scope of this paper. Moreover, within an observation cycle, the drift-parameters are assumed to be constant. If the drift parameters change frequently, our proposed framework must be used in conjunction with periodic system re-identifications.

D. Deterministic signal model

Often in the sensor network literature, the signal of interest is assumed to be constant over the observation duration, e.g., [2],[3]. This means $x_n = \beta, \forall n$. However, such a signal model may be too simplistic for real applications and we consider a generalization of the form,

$$x_n = \sum_{p=0}^{P} \beta_p n^p, \qquad \beta_p \in \mathbf{R}, \qquad (5)$$

where β_p-s are the unknown constants that need to be estimated and P is the order of the polynomial time-series that is assumed to be known. In vector notations, the polynomial signal $\boldsymbol{x} \triangleq [x_1, x_2, \ldots, x_N]'$ assumes a linear form $\boldsymbol{x} = \boldsymbol{X}\boldsymbol{\beta}$, where $\boldsymbol{X}$ is the Vandermonde matrix

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2^P \\ \vdots & \vdots & \ddots & \vdots \\ 1 & N & \cdots & N^P \end{bmatrix}, \text{ and } \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_P \end{bmatrix}. \qquad (6)$$

It may be noted that time-varying signals in different applications are approximated by either polynomials or piecewise-polynomials [20],[21]. For example, a polynomial regression-based data gathering algorithm for environmental monitoring applications was suggested in [22]. Also, a polynomial spline approximation of stationary random processes was applied to Clarke's model of multipath fading channels in [23].

This completes the description of the signal and observation noise considered in this paper. We intend to derive the closed form CRB (for the estimation of β_p-s) in terms of the signal (P, N) and noise $\{\sigma_m^2, \rho_m, \gamma_m\}$ parameters. This would help us characterize the performance of a sensor network and increase our understanding about its estimation capabilities.

E. Related work

A related area of work is the study of systematic-bias or model-error estimation schemes using multiple, and sometimes collaborative sensors. In the radar signal processing literature, the process of model-based estimation and subsequent removal of systematic errors prior to target tracking is known as sensor-registration [24], [25]. In the weather research literature, the serially-correlated forecasting error arising due to modeling

deficiency is often considered separately and tracked along-side the model parameters [26], [27]. In the sensor network literature, drift-aware networks perform learning-based collaborative bias estimation to enhance the effective lifetime of the network [9], [28]. However, in this paper, we are focused on the quality of estimation in the presence of systematic errors, rather than techniques on mitigating systematic errors.

Several researchers have studied the Cramér-Rao bounds for polynomial (or polyphase) signal estimation in the presence of independent (or correlated) noise. The CRB is usually obtained from the inverse of the Fisher Information Matrix (FIM) [10]. For Additive White Gaussian Noise (AWGN), the large sample approximation of FIM is known to be a multiple of the Hilbert Matrix (e.g., [29]). The second order approximation was derived in [30] in the context of polynomial phase signals. For a mixture of additive and multiplicative white Gaussian noise, the large sample FIM was shown in [31] to be a scalar multiple of the AWGN case. *Our primary contribution in this paper is the derivation of (approximate) closed-form CRB for polynomial signal estimation in the presence of a mixture of white (measurement noise) and AR(1) Gaussian noise (drift).* We also discuss the non-stationary case when the autoregressive parameter is equal to 1. To the knowledge of the authors, polynomial signal estimation in such a mixture of noise has not been considered earlier. As mentioned earlier, the study of estimating polynomial signals in AR(1)+White noise would help us characterize the capability of a sensor network to infer the parameters of a time-varying signal using sensors with drift.

The rest of the paper is organized as follows. In Section II, we formally list the assumptions and set up the problem for two scenarios based on the initial state of the sensor. In Section III, we consider the single-sensor case and derive large sample approximations of the CRB for $\rho < 1$ and $\rho = 1$. In Section IV, we extend the results to multiple sensors having different drift characteristics. In Section V, we demonstrate the application of the results to a bandwidth limited sensor network that communicates only quantized observations. Concluding remarks and scope of future research are provided in Section VI.

II. PROBLEM FORMULATION

To be more specific about our problem framework, we formally state the assumptions here. *1) Same phenomenon:* Each sensor is observing the same physical phenomenon (e.g., temperature), which is modeled as a time-polynomial signal. In our framework, if multiple signals are to be sensed (e.g., temperature and humidity), the observations have to be transmitted separately and the parameters independently estimated. *2) Known drift statistics:* The drift in each sensor is modeled as AR(1) Gaussian time-series with known statistics - namely the autoregressive and strength parameters, ρ_m and γ_m respectively. *3) Spatially uncorrelated noise and drift:* We assume that the observations at all the sensors at a particular instant are independent, conditioned on the signal magnitude at that instant. In other words, though the observation noise samples $d_{n,m} + w_{n,m}$ are temporally correlated (due to drift),

there is no spatial correlation among them. *4) Synchronized observations:* The clocks of the sensors are synchronized and they have identical sampling intervals. *5) Parallel sensor network with fusion center:* We assume that the sensors do not collaborate among themselves and rather communicate their observations only to the sink. *6) Reliable signal transmission:* In Sections III and IV, we assume that the noisy observations are communicated perfectly to the sink without any further distortion. We call this the *full-precision* case which helps obtain a benchmark performance. However, there may be cases when, due to power and bandwidth limitations at the sensor nodes, reliable communication of full-precision observations may be impossible. In such cases, a digital communication based framework in conjunction with efficient channel coding can be used to reliably transmit only a finite number of bits [2], [32]. In Section V, we will discuss inference using digitized observations where all sensor nodes perform quantization with identical fidelity[1]. With the aforementioned assumptions, our goal in this paper is to derive approximate closed form expressions for the Cramér-Rao bounds.

A. Sensor calibration and noise covariance

Consider a single sensor with noise and drift properties denoted by $\{\sigma^2, \gamma, \rho\}$ (the subscript m is dropped for most of this subsection). Assume that inference has to be performed from noisy observations z_n at instants $n = 1, \ldots, N$. Let $\tau > 0$ denote the time-instants elapsed since the sensor was last calibrated, i.e., when the drift component was set to zero by correcting the baseline. Since the duration of inference starts from instant 1, it follows that the drift sequence was set to zero at instant $1 - \tau$ (which is a non-positive index, a slight notational inconvenience). Therefore, the drift sequence proceeds as follows,

$$d_{1-\tau} = 0, \; d_{2-\tau} = \delta_1, \; d_{3-\tau} = \rho\delta_1 + \delta_2 \ldots \text{etc..} \quad (7)$$

Since each of the drift innovations $\delta_i \sim \mathcal{N}(0, \gamma\sigma^2)$, we have

$$\text{Var}(d_n) = \gamma\sigma^2(1 + \rho^2 + \cdots + \rho^{2(\tau+n-1)}) =: \gamma\sigma^2 S_n^\tau, \quad (8)$$

where the summation within the parenthesis is defined as S_n^τ. Though the drift sequence $\{d_n\}_{n=1}^N$ is stationary for large magnitudes of τ,

$$\lim_{\tau \to \infty} S_n^\tau = \frac{1}{1 - \rho^2}, \; \rho < 1, \quad (9)$$

it is not stationary in the transient stage, for which,

$$S_1^\tau < S_2^\tau < \cdots < S_N^\tau, \quad \tau < \infty. \quad (10)$$

In this paper, though we would solve the estimation problem for a general τ, we would refer to the limiting cases $\tau = 1$ as *calibrated* (C) and $\tau \to \infty$ as *uncalibrated* (U) respectively.

---

[1]For sensor networks where there is a constraint on system-wide bandwidth (summed across all nodes), different sensors nodes may be assigned different fidelity of quantization based on their observation quality, e.g., [33]. Though outside the scope of this paper, optimal fidelity assignment for sensors in the presence of drift is a challenging topic worthy of future research.

We define $\boldsymbol{R}$ such that $\gamma\sigma^2\boldsymbol{R}$ is the covariance matrix of the drift vector $\boldsymbol{d} \triangleq [d_1, d_2, \ldots, d_N]'$

$$\mathbb{E}[\boldsymbol{d}'\boldsymbol{d}] = \gamma\sigma^2 \begin{bmatrix} S_1^\tau & \rho S_1^\tau & \rho^2 S_1^\tau & \cdots & \rho^{N-1} S_1^\tau \\ \rho S_1^\tau & S_2^\tau & \rho S_2^\tau & \cdots & \rho^{N-2} S_2^\tau \\ \rho^2 S_1^\tau & \rho S_2^\tau & S_3^\tau & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \rho^{N-1} S_1^\tau & \rho^{N-2} S_2^\tau & \cdots & \cdots & S_N^\tau \end{bmatrix}$$
$$=: \gamma\sigma^2 \boldsymbol{R}. \tag{11}$$

For the uncalibrated case, the diagonal elements are the same ($S_n^\tau = \frac{1}{1-\rho^2}$), and $\boldsymbol{R}$ takes the shape of the well known stationary matrix (e.g., [34]),

$$\mathbb{E}[\boldsymbol{d}'\boldsymbol{d}] = \frac{\gamma\sigma^2}{1-\rho^2} \begin{bmatrix} 1 & \rho & \cdots & \rho^{N-1} \\ \rho & 1 & \cdots & \rho^{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \cdots & 1 \end{bmatrix}, \tag{12}$$

sometimes referred to as the Kac–Murdock–Szegö matrix.

For the mth sensor, we denote the covariance of the drift sequence by $\boldsymbol{R}_m$. Let $\boldsymbol{\Sigma}_m$ denote the covariance matrix of the total error (AR(1)+White noise), i.e., $\boldsymbol{\epsilon}_m = \boldsymbol{w}_m + \boldsymbol{d}_m$, so that,

$$\boldsymbol{\Sigma}_m = \sigma_m^2 \left(\boldsymbol{I} + \gamma_m \boldsymbol{R}_m \right), \tag{13}$$

where $\boldsymbol{I}$ is the $N \times N$ identity matrix.

B. Maximum-Likelihood estimation and Cramér-Rao bound

Given the linear observation model (1) and noise covariance described in Section II-A, the Maximum-Likelihood (ML) estimator of $\boldsymbol{\beta}$ at the sink is of the form (for a reference on estimation theory, see [10], [11]),

$$\widehat{\boldsymbol{\beta}}^{\text{ML}} = \boldsymbol{J}^{-1} \boldsymbol{X}' \sum_{m=1}^{M} \boldsymbol{\Sigma}_m^{-1} \boldsymbol{z}_m, \tag{14}$$

where $\boldsymbol{J}$ is known as the Fisher Information matrix,

$$\boldsymbol{J} \triangleq \boldsymbol{X}' \left(\sum_{m=1}^{M} \boldsymbol{\Sigma}_m^{-1} \right) \boldsymbol{X}, \tag{15}$$

It is well known in estimation theory [10] that within the class of unbiased estimators, the ML estimator of a linear model is optimal in terms of estimation variance [10]. Also, the least possible estimation variance is provided by the Cramér-Rao lower bound, $\boldsymbol{V} \triangleq \boldsymbol{J}^{-1}$, so that

$$\mathbb{E}\left[(\beta_p - \widehat{\beta}_p^{\text{ML}})^2 \right] \geq \boldsymbol{V}_{p,p} = \left[\boldsymbol{J}^{-1} \right]_{p,p}, \tag{16}$$

for $0 \leq p \leq P$. Since Equation (16) holds with a strict equality for linear models, the CRB is an appropriate performance metric for our problem.

It is unclear from (15) and (16) exactly how the estimation variance $\boldsymbol{V}_{p,p}$ depends on the drift parameters $\sigma_m^2, \gamma_m, \rho_m$, signal parameters P and the sample size N. *Our goal in this paper is to derive large sample approximations for these Cramér-Rao bounds and thereby provide insight into the behavior of estimation performance as it relates to the drift properties of the sensors.*

III. MAIN RESULT: SINGLE SENSOR CASE

Towards obtaining the CRB for the general case of multiple sensors, we first consider the single-sensor scenario and thereby obtain the core results of this paper. Subsequent applications of these core results to the multi-sensor scenario (Section IV) and bandwidth limited networks (Section V), as we shall see later, will be somewhat straightforward extensions.

While analyzing the single-sensor scenario, for notational brevity, we drop the sensor-index subscript m from σ_m^2, ρ_m, γ_m, $\boldsymbol{R}_m$ and $\boldsymbol{\Sigma}_m$ for the remainder of this section. Note from (15) and (16) that, for a single sensor, the Fisher Information matrix is $\boldsymbol{J} = \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}$ and the Cramér-Rao bound is $[\boldsymbol{V}]_{p,p} = [\boldsymbol{J}^{-1}]_{p,p}$. In the following discussion, we first compute the inverses of the disturbance covariance matrices. Next, we use those results to derive the CRB.

A. Inverse of disturbance covariance

Our goal is to approximate $\boldsymbol{\Sigma}^{-1}$ in a form that are analytically tractable. The exact closed form expression for $\boldsymbol{\Sigma}_U^{-1}$ (for $\tau \to \infty$) is known to be quite complicated (e.g., p-53 of [35], [36]) and we will not use it. The authors have not found a closed form expression for either $\boldsymbol{\Sigma}_C^{-1}$ (the $\tau = 1$ case) or $\boldsymbol{\Sigma}^{-1}$ (for general τ) in the literature.

In this paper, we propose an approximation to $\boldsymbol{\Sigma}^{-1}$ that is novel to the best of knowledge of the authors. We suggest the following form for the inverse

$$(\boldsymbol{I} + \gamma\boldsymbol{R})^{-1} = \boldsymbol{I} - \nu\boldsymbol{M} + \mathcal{O}(y^N), \tag{17}$$

where ν is a constant, $\boldsymbol{M}$ is a matrix with certain structure (to be described later), and y is a quantity less than 1, so that $\mathcal{O}(y^N)$ represents a term that vanishes exponentially with sample size N. The idea is that, for large N, we should be able to use the approximation

$$(\boldsymbol{I} + \gamma\boldsymbol{R})^{-1} \approx \boldsymbol{I} - \nu\boldsymbol{M} \tag{18}$$

towards computing the CRB. The specifics of this approximation are laid out in Proposition 1.

Proposition 1: Based on drift parameters γ and ρ, define the following constants,

$$y \triangleq \frac{1}{2} \left[\frac{\gamma+1}{\rho} + \rho - \sqrt{\left(\frac{\gamma+1}{\rho} + \rho \right)^2 - 4} \right],$$
$$\nu \triangleq \frac{y\gamma}{\rho(1-y^2)}, \text{ and } \kappa \triangleq \frac{y(\rho-y)}{1-\rho y}. \tag{19}$$

Then $(\boldsymbol{I} + \gamma\boldsymbol{R})^{-1} = \boldsymbol{I} - \nu\boldsymbol{M} + \mathcal{O}(y^N)$, where $\boldsymbol{M}$ has the structure shown in Equation (21), with

$$a_i \triangleq 1 + y^{2(i-1)}\eta_\tau, \; b_j \triangleq 1 + y^{2(j-1)}\kappa,$$
$$\eta_\tau \triangleq \frac{1-y^2}{1-\rho y + \varrho_\tau y/\rho} - 1, \; \varrho_\tau \triangleq \frac{\rho^{2\tau}}{1 + \rho^2 + \cdots + \rho^{2\tau-2}}. \tag{20}$$

Proof: See Appendix A. ∎

The constant y can perhaps be more conveniently described as the smaller of the roots of the quadratic equation (both

$$\boldsymbol{M} = \begin{bmatrix} a_1 & ya_1 & y^2a_1 & \cdots & \cdots & \cdots & y^{N-2}a_1 & y^{N-1}b_1 \\ ya_1 & a_2 & ya_2 & \cdots & \cdots & y^{N-4}a_2 & y^{N-3}b_2 & \vdots \\ y^2a_1 & ya_2 & a_3 & & y^{N-5}b_3 & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \cdot^{\cdot^{\cdot}} & \vdots & \vdots & \vdots \\ \vdots & \vdots & & \cdot^{\cdot^{\cdot}} & \ddots & \vdots & \vdots & \vdots \\ \vdots & y^{N-4}a_2 & y^{N-5}b_3 & \cdots & \cdots & b_3 & yb_2 & y^2b_1 \\ y^{N-2}a_1 & y^{N-3}b_2 & \cdots & \cdots & \cdots & yb_2 & b_2 & yb_1 \\ y^{N-1}b_1 & \cdots & \cdots & \cdots & \cdots & y^2b_1 & yb_1 & b_1 \end{bmatrix}, \quad \begin{aligned} a_i &= 1 + y^{2(i-1)}\eta_\tau, \\ b_j &= 1 + y^{2(j-1)}\kappa. \end{aligned} \quad (21)$$

of which are positive) $y^2 - \left(\frac{\gamma+1}{\rho} + \rho\right) + 1 = 0$. It is easy to establish that $y < 1$ for $\rho \le 1$ and for all γ, so that $y^N \to 0$ for large sample size N. Also note that for the two special cases of $\tau = 1$ and $\tau = \infty$, $\varrho_C = \rho^2, \eta_C = -y^2$ and $\varrho_U = 0, \eta_U = \kappa$.

Remark 1.1: Identity: The following identity will be used in several places in this paper and follows from definition (19) after some algebraic manipulations,

$$\gamma y = (\rho - y)(1 - \rho y). \quad (22)$$

Remark 1.2: It may be noted that $\boldsymbol{M}$ is defined entirely by the diagonal elements $a_1, a_2, \ldots, a_{\lfloor\frac{N}{2}\rfloor}$ (counted from top) and $b_1, b_2, \ldots, b_{\lceil\frac{N}{2}\rceil}$ (counted from bottom). From the expressions for a_i and b_j in Equation (21), we note that for large N, both $a_{\lfloor\frac{N}{2}\rfloor}$ and $b_{\lceil\frac{N}{2}\rceil}$ converge to the same value, namely 1. That means that it should not matter whether the anti-diagonal elements are expressed in terms of a_i-s or b_j-s. In Proposition 1, we describe the antidiagonal elements in terms of b_j-s just for the sake of being definitive.

B. Computation of Fisher Information Matrix

We use Proposition 1 in (13) and (15) to compute the Fisher Information matrix,

$$\boldsymbol{J} \approx \frac{1}{\sigma^2}\boldsymbol{X}'(\boldsymbol{I} - \nu\boldsymbol{M})\boldsymbol{X}, \quad (23)$$

which would be inverted later to obtain the CRB-s. We consider the matrices $\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{X}'\boldsymbol{M}\boldsymbol{X}$ individually before summing them up. We would approximate all the elements of these matrices as polynomials in N (i.e., the sample size). This will help derive approximations of the CRB that are correct upto the second order for $\rho \le 1$.

We note from (6) that $\boldsymbol{X}'\boldsymbol{X}$ is a Hankel (equal skew-diagonal elements) matrix,

$$[\boldsymbol{X}'\boldsymbol{X}]_{k,l} = \sum_{n=1}^{N} n^{k+l}, \quad 0 \le k, l \le P. \quad (24)$$

It is a well known result, e.g., [30], [31] that summations of the form (24) can be written as

$$\sum_{n=1}^{N} n^q = \sum_{i=0}^{q} \mathcal{B}_{q,i} N^{q+1-i}, \quad q \ge 0, \quad \text{where} \quad (25)$$

$$\mathcal{B}_{q,0} \triangleq \frac{1}{q+1}, \mathcal{B}_{q,1} \triangleq \frac{1}{2}, \mathcal{B}_{q,2} \triangleq \frac{q}{12}, \mathcal{B}_{q,3} \triangleq 0, \text{etc.}$$

A general form for $\mathcal{B}_{q,i}$ can be found, for example, on p-1, [37]. However, all the results in this paper will be established using the first four terms which we have enumerated in (25). Similar polynomial expressions can be obtained for $\boldsymbol{X}'\boldsymbol{M}\boldsymbol{X}$.

Proposition 2: For $0 \le k, l \le P$, we have

$$[\boldsymbol{X}'\boldsymbol{M}\boldsymbol{X}]_{k,l} = \sum_{i=0}^{k+l} \mathcal{A}_{k,l,i} N^{k+l+1-i} + \alpha_{k,l}^{(\tau)} + \mathcal{O}(N^{k+l+1}y^N),$$

$$(26)$$

where some leading constants $\mathcal{A}_{k,l,i}$ and $\alpha_{k,l}^{(\tau)}$ are

$$\mathcal{A}_{k,l,0} \triangleq \frac{1+y}{1-y}\frac{1}{k+l+1}, \quad \forall k, l,$$

$$\mathcal{A}_{k,l,1} \triangleq \frac{1+2\kappa-2y-y^2}{2(1-y)^2}, \quad k+l \ge 1, \quad (27)$$

$$\alpha_{0,0}^{(\tau)} \triangleq \frac{-2y+\eta_\tau+\kappa}{(1-y)^2}, \quad \text{etc.}$$

Proof: See Appendix B. ∎

The exact form of $\mathcal{A}_{k,l,i}$ and $\alpha_{k,l}^{(\tau)}$ are provided in Appendix B. From (26), it is clear that the Fisher information for calibrated and uncalibrated cases differ only by a constant (dependence on τ is explicitly indicated). This means that when sample size is large, estimation accuracy will be similar for both the cases. We will elaborate this point later. Next, we use the result in Proposition 2 to derive the CRBs for the cases when $\rho < 1$ and $\rho = 1$.

C. Cramér-Rao Bounds for $\rho < 1$

Using (24),(25) and (26) in (23), we obtain the compact second order expression for the Fisher Information Matrix,

$$\boldsymbol{J} = N\boldsymbol{E}\left[\xi_0\boldsymbol{H} + \frac{\xi_1\boldsymbol{ee}' + \xi_2^{(\tau)}\boldsymbol{ff}'}{N} + \mathcal{O}\left(\frac{1}{N^2}\right)\right]\boldsymbol{E}, \quad (28)$$

where $\boldsymbol{E}, \boldsymbol{H}$ are $(P+1) \times (P+1)$ matrices, $\boldsymbol{e}, \boldsymbol{f}$ are $(P+1)$-dimensional vectors and $\xi_0, \xi_1, \xi_2^{(\tau)}$ are constants defined by

$$\boldsymbol{E} \triangleq \text{diag}\{1, \ldots, N^P\}, \quad \boldsymbol{H}_{k,l} \triangleq \frac{1}{k+l+1},$$

$$\boldsymbol{e} \triangleq [1, \ldots, 1]', \quad \boldsymbol{f} \triangleq [1, 0, \ldots, 0]',$$

$$\xi_0 \triangleq \frac{1}{\sigma^2}\left[1 - \nu\frac{1+y}{1-y}\right], \xi_1 \triangleq \frac{1}{\sigma^2}\left[\frac{1}{2} - \nu\mathcal{A}_1\right], \quad (29)$$

$$\xi_2^{(\tau)} \triangleq -\xi_1 - \frac{1}{\sigma^2}\nu\alpha_{0,0}^{(\tau)},$$

where $0 \leq k, l \leq P$. The notation $\mathcal{A}_1$ is actually $\mathcal{A}_{k,l,1}$ with the subscripts dropped, since $\mathcal{A}_{k,l,1}$ does not depend on k and l (see (27)). $\boldsymbol{H}$ is the well known Hilbert matrix [30].

The leading constant ξ_0 can be simplified further based on definitions in (19) and the identity (22),

$$\xi_0 = \frac{1}{\sigma^2} \left[1 + \frac{\gamma}{(1-\rho)^2} \right]^{-1}. \qquad (30)$$

The CRB will be obtained as the inverse of the FIM described in (28). Since $\rho < 1$, we ensure from (30) that $\xi_0 > 0$. Therefore, the FIM of the form (28) can be inverted assuming $\xi_0 \boldsymbol{H}$ as the dominant term. Such an inversion was performed in [30] in the context of polynomial-phase signals. We summarize the analysis in [30] in the form of Lemma 1.

Lemma 1: [30]: Let $\boldsymbol{H}$ be the Hilbert matrix and $\boldsymbol{e}, \boldsymbol{f}$ vectors as defined in (29). Let c_0, c_1, c_2 be constants such that $c_0 \neq 0$. Then, for $0 \leq p \leq P$, the diagonal elements are

$$\left[\left[c_0 \boldsymbol{H} + \frac{1}{N} \left(c_1 \boldsymbol{e} \boldsymbol{e}' + c_2 \boldsymbol{f} \boldsymbol{f}' \right) \right]^{-1} \right]_{p,p} =$$
$$\frac{K_{P,p}}{c_0} \left[\frac{1}{2p+1} - \frac{1}{Nc_0} \left(c_1 + c_2 L_{P,p} \right) + \mathcal{O}\left(\frac{1}{N^2} \right) \right], \qquad (31)$$

where $K_{P,p}$ and $L_{P,p}$ are defined by

$$K_{P,p} \triangleq \left[(P+p+1) \binom{P+p}{p} \binom{P}{p} \right]^2, \quad L_{P,p} \triangleq \left[\frac{P+1}{p+1} \right]^2. \qquad (32)$$

The idea behind Lemma 1 is that, for sufficiently large N, terms of order $\mathcal{O}(\frac{1}{N^2})$ in (31) can be ignored and we obtain approximate closed form expressions for the diagonal elements. The approximation in (31) is accurate only for small relative magnitude of the second order term, i.e., small magnitudes of $\frac{(2p+1)(c_1+c_2 L_{P,p})}{c_0 N}$. Lemma 1 can be used now to invert (28). The Cramér-Rao bound (diagonal terms) obtained in such a manner are summarized in Proposition 3.

Proposition 3: For $\rho < 1$ and sufficiently large N, the CRB for estimating β_p is

$$[\boldsymbol{V}]_{p,p} \approx \frac{K_{P,p}}{N^{2p+1}\xi_0} \left[\frac{1}{2p+1} - \frac{1}{N\xi_0} \left(\xi_1 + \xi_2^{(\tau)} L_{P,p} \right) \right], \qquad (33)$$

for $0 \leq p \leq P$.

A few remarks due to Proposition 3 are in order.

Remark 3.1: Asymptotic performance: As expected, the estimation performance does not depend on transient information (τ) when the number of samples used for inference is sufficiently large. From (30) and (33), the first order approximation of CRB is

$$[\boldsymbol{V}]_{p,p} \approx \left[1 + \frac{\gamma}{(1-\rho)^2} \right] \frac{\sigma^2 K_{P,p}}{N^{2p+1}(2p+1)}, \qquad (34)$$

which is true for any value of τ.

Remark 3.2: Equivalent AWGN noise: For estimation of polynomial signals, the effect of drift is asymptotically (upto first order) equivalent to scaling the measurement noise by a factor of $1 + \gamma/(1-\rho)^2$, where γ, ρ characterizes the drift properties of a sensor.

Remark 3.3: Constant signal: $x_n = \beta_0, \forall n$: Simpler and more precise expressions for CRB may be obtained for a constant signal. When $P = 0$, we can directly proceed from (28). For this case, we have $\boldsymbol{e}\boldsymbol{e}' = \boldsymbol{f}\boldsymbol{f}' = \boldsymbol{E} = \boldsymbol{H} = 1$, hence

$$\boldsymbol{V} \approx \frac{1}{\xi_0 N + \xi_1 + \xi_2^{(\tau)}} = \frac{1}{\xi_0 N - \nu\alpha_{0,0}^{(\tau)}/\sigma^2}, \qquad (35)$$

where the scalar $\boldsymbol{V}$ denotes the Cramér-Rao bound.

Remark 3.4: Dependence on τ: In terms of estimation of a signal, intuitively, the transient state $\tau < \infty$ should be more informative, since the drift starts from a known point, rather than an unknown point. This intuition is corroborated by equation (33), from which it is easy to derive that $[\boldsymbol{V}]_{p,p}$ increases with τ (shown below). In conjunction with Remark 3.1, we can conclude that, though calibration always results in better estimates, the relative gain in performance $([\boldsymbol{V}_U]_{p,p} - [\boldsymbol{V}_C]_{p,p})/[\boldsymbol{V}_U]_{p,p}$) diminishes with sample size.

Proof that $[\boldsymbol{V}]_{p,p}$ increases with τ: Since $\xi_0 > 0$ (from (30)), it suffices to show that $\xi_2^{(\tau)}$ decreases with τ (see (33)), or equivalently, $\alpha_{0,0}^{(\tau)}$ increases with τ (see (29), note that $\nu > 0$), which is further equivalent to showing that η_τ increases with τ (see (27)). The last condition is verified from (20), since ρ_τ is a decreasing function of τ and $0 < \rho, y < 1$.

Remark 3.5: Approximation region: The CRB as expressed by (33) is accurate only for cases when the relative magnitudes of the second order terms,

$$\epsilon_p = \frac{(2p+1)(\xi_1 + \xi_2^{(\tau)} L_{P,p})}{\xi_0 N}, \quad 0 \leq p \leq P, \qquad (36)$$

are small. This condition helps specify the operating values of $\sigma^2, \gamma, \rho, P$ and N for which the closed-form CRB (33) can be used for performance analysis.

D. Cramér-Rao Bounds for $\rho = 1$

For the case when the AR parameter $\rho = 1$, the drift phenomenon is a non-stationary even in the limit of $\tau \to \infty$, since it is a random walk (e.g., [9]) where the bias builds up with time and is unbounded. Hence, the CRB for the uncalibrated scenario is infinity and we consider only the transient case $\tau < \infty$ here.

For $\rho = 1$, the leading term of the Fisher information matrix $\boldsymbol{J}$ (28) vanishes since (30) implies that $\xi_0 = 0$. Hence, terms of order $\mathcal{O}(N^{-2})$ need to be considered. In particular, using definitions (19), (27) and (29), the constants can further be simplified as

$$\xi_0 = 0, \; \xi_1 = 0, \; \varrho_\tau = \frac{1}{\tau}, \; \widetilde{\gamma} \triangleq \sqrt{1 + 4\gamma^{-1}},$$
$$\xi_2^{(\tau)} = \frac{1}{\sigma^2} \left[\frac{2}{\widetilde{\gamma}+1} \left(1 + \frac{2\tau}{\widetilde{\gamma}-1} \right) \right]^{-1}, \qquad (37)$$

and we subsequently note (see Appendix C) that the FIM is of the structure depicted in equation (38),

$$\boldsymbol{J} = \boldsymbol{E} \begin{bmatrix} \xi_2^{(\tau)} & \frac{1}{N}\xi_3^{(\tau)} \boldsymbol{f}'_P + \mathcal{O}\left(\frac{1}{N^2}\right) \\ * & \frac{1}{N} \boldsymbol{D} \left[\xi_4 \boldsymbol{H}_P + \frac{1}{N}\boldsymbol{F} + \mathcal{O}\left(\frac{1}{N^2}\right) \right] \boldsymbol{D} \end{bmatrix} \boldsymbol{E},$$
$$\text{with } \boldsymbol{F} = \xi_5 \boldsymbol{e}_P \boldsymbol{e}'_P + \xi_6^{(\tau)} \boldsymbol{f}_P \boldsymbol{f}'_P, \qquad (38)$$

with the matrix $\boldsymbol{D}$ and constants $\xi_3^{(\tau)}, \xi_4, \xi_5, \xi_6^{(\tau)}$ defined as

$$\boldsymbol{D} \triangleq \mathrm{diag}\{1, \ldots, P\}, \ \xi_3^{(\tau)} \triangleq -\frac{1}{\sigma^2}\nu\alpha_{1,0}^{(\tau)}, \ \xi_4 \triangleq \frac{1}{\sigma^2\gamma},$$
$$\xi_5 \triangleq -\frac{1}{\sigma^2}\frac{y^2}{(1-y)^3}, \ \xi_6^{(\tau)} \triangleq -\xi_5 - \frac{1}{\sigma^2}\nu\alpha_{1,1}^{(\tau)}, \quad (39)$$

and $\boldsymbol{e}_P, \boldsymbol{f}_P$ and $\boldsymbol{H}_P$ being the P-dimensional equivalents of $\boldsymbol{e}, \boldsymbol{f}$ and $\boldsymbol{H}$ (defined in (29)) respectively. The CRB is obtained from the FIM (38) by using block-inversion and Lemma 1 (see Appendix C).

Proposition 4: For $\rho = 1$ and sufficiently large N, the CRB for estimating β_p, $[\boldsymbol{V}]_{p,p}$ is equal to

$$[\boldsymbol{V}]_{p,p} \approx \begin{cases} \frac{1}{\xi_2^{(\tau)}}\left[1 + \frac{1}{N}\frac{P^2\left(\xi_3^{(\tau)}\right)^2}{\xi_2^{(\tau)}\xi_4}\right], & p = 0, \\ \frac{K_{P-1,p-1}}{N^{2p-1}p^2\xi_4}\left[\frac{1}{2p-1} - \frac{1}{N}\widetilde{\xi}\right], & 1 \le p \le P, \end{cases}$$
$$\text{with } \widetilde{\xi} = \frac{1}{\xi_4}\left(\xi_5 + \left(\xi_6^{(\tau)} - \frac{\left(\xi_3^{(\tau)}\right)^2}{\xi_2^{(\tau)}}\right)\frac{P^2}{p^2}\right). \quad (40)$$

A few remarks due to Proposition 4 are noted below.

Remark 4.1: Constant Signal: $x_n = \beta_0, \forall n$: For a random walk drift, a constant signal can only be estimated inconsistently with asymptotic variance

$$\boldsymbol{V} \approx \frac{1}{\xi_2^{(\tau)}} = \frac{2\sigma^2}{\widetilde{\gamma}+1}\left(1 + \frac{2\tau}{\widetilde{\gamma}-1}\right). \quad (41)$$

It is explicitly seen from (41) that the variance increases with white noise variance σ^2, the drift strength γ (note that $\widetilde{\gamma}$ decreases with γ, see (37)), and the time since last calibration, τ. This phenomenon of inconsistent estimation can be intuitively understood as follows. Even without the white noise component in (3), the observation never captures independent readings of β_0. Rather, the samples are

$$z_1 = \beta_0 + d_\tau + w_1,$$
$$z_2 = \beta_0 + d_{\tau+1} + w_2 = (\beta_0 + d_\tau) + \delta_{\tau+1} + w_2, \text{ etc.,} \quad (42)$$

which means that $\beta_0 + d_\tau$ appears together in all subsequent observations, thereby making β_0 indistinguishable from d_τ. Since d_τ has a finite variance, β_0 can only be estimated inconsistently.

Remark 4.2: Time Features: For other parameters whose effect on the signal vary with time, i.e., β_p for $p \ge 1$, the CRB is (to first order) equivalent to the case of estimating the derivative of the signal in drift-only noise. We note that the derivative of the signal is $x'(t) = \sum_{p=1}^{P} p\beta_p t^{p-1}$ and the forward difference (see (3)),

$$z_{n+1} - z_n = x_{n+1} - x_n + d_{n+1} - d_n + w_{n+1} - w_n$$
$$= x'(n)(\Delta T) + \delta_n + w_{n+1} - w_n, \quad (43)$$

which is an estimator of the derivative, contains independent drift innovations δ_n with variance $\gamma\sigma^2$. Hence in Proposition 4, the variance of estimating β_p is scaled down by a factor of p^2 compared to the equivalent case with white noise $\gamma\sigma^2$.

Remark 4.3: Approximation region: The CRB as expressed by (40) is accurate only for cases when the relative magnitude of the second order terms,

$$\epsilon_p = \begin{cases} \frac{1}{N}\frac{P^2\left(\xi_3^{(\tau)}\right)^2}{\xi_2^{(\tau)}\xi_4}, & p = 0, \\ \frac{2p-1}{N}\widetilde{\xi}, & 1 \le p \le P. \end{cases} \quad (44)$$

are small. This condition helps specify the operating values of σ^2, γ, P and N for which the closed-form CRB (40) can be used for performance analysis.

E. Numerical Results - Maximum Relative Error

In the remainder of this section, we demonstrate the accuracy of the closed-form performance approximations using representative examples. We shall consider the approximations expressed in both the forms of Fisher Infomation matrix (Equations (28) and (38), with $\mathcal{O}\left(\frac{1}{N^2}\right)$ terms truncated) and Cramér-Rao bounds (Equations (33) and (40)).

We use the metric called Maximum Relative Error (MRE) which was also used in [30]. This involves computing both the exact (16) and approximate CRB-s. The relative deviations of the approximations are then calculated for all $0 \le p \le P$ and the largest one is called the MRE. We denote *1)* $[\boldsymbol{V}^{\mathrm{TH}}]_{p,p}$ as the theoretical CRB as in (16) and *2)* $[\boldsymbol{V}^{\mathrm{AP}}]_{p,p}$ as the approximate CRB derived either through *2a) FIM approx:* inversion of the intermediate FIM, i.e., $[\boldsymbol{V}^{\mathrm{AP}}]_{p,p} = [\boldsymbol{J}^{-1}]_{p,p}$, in Equations (28) and (38) with $\mathcal{O}\left(\frac{1}{N^2}\right)$ terms truncated, or *2b) CRB approx:* final Cramér-Rao bounds in Equations (33) and (40). Then the maximum relative error is defined by

$$\mathrm{MRE} = \max_{p=0,1,\cdots,P} \frac{|[\boldsymbol{V}^{\mathrm{TH}}]_{p,p} - [\boldsymbol{V}^{\mathrm{AP}}]_{p,p}|}{[\boldsymbol{V}^{\mathrm{TH}}]_{p,p}}. \quad (45)$$

In short, the MRE summarizes the approximation error over all components of the parameter vector.

As an example, we consider the estimation of a cubic-polynomial signal, i.e., $P = 2$. Since our performance bounds are asymptotically accurate, it is expected that the MRE will decrease with increasing sample size N. In Figure 1, we display the sample size N_ϵ required for 95% accuracy, or in other words, MRE $< \epsilon = 0.05$. Since MRE is a ratio of variances, it does not depend on the measurement noise variance σ^2. While considering the $\rho < 1$ case, we have displayed the wide parameter region $\gamma \in [10^{-3}, 1]$, $\rho \in [0.6, 0.97]$ in Figures 1(a) and 1(b) - which demonstrate the uncalibrated ($\tau \to \infty$) and calibrated ($\tau = 1$) scenarios respectively. The $\rho = 1, \tau = 1$ case is demonstrated in 1(c), where we have displayed the parameter region $\gamma \in [10^{-3}, 1]$.

Figure 1 provides useful guidance on the applicability and limitations of the performance bounds derived in this paper. Firstly, the performance bounds are found to be reasonably accurate over a wide range of possible parameter values and for a moderate number (tens and hundreds) of samples. For accurate performance prediction in stationary drift (Figures 1(a) and 1(b)), higher values of drift-autocorrelation (ρ) and drift-strength (γ) generally requires larger observation durations (N_ϵ). This is predicted by the approximation-region condition in (36), since the denominator ξ_0 (see (30)) is
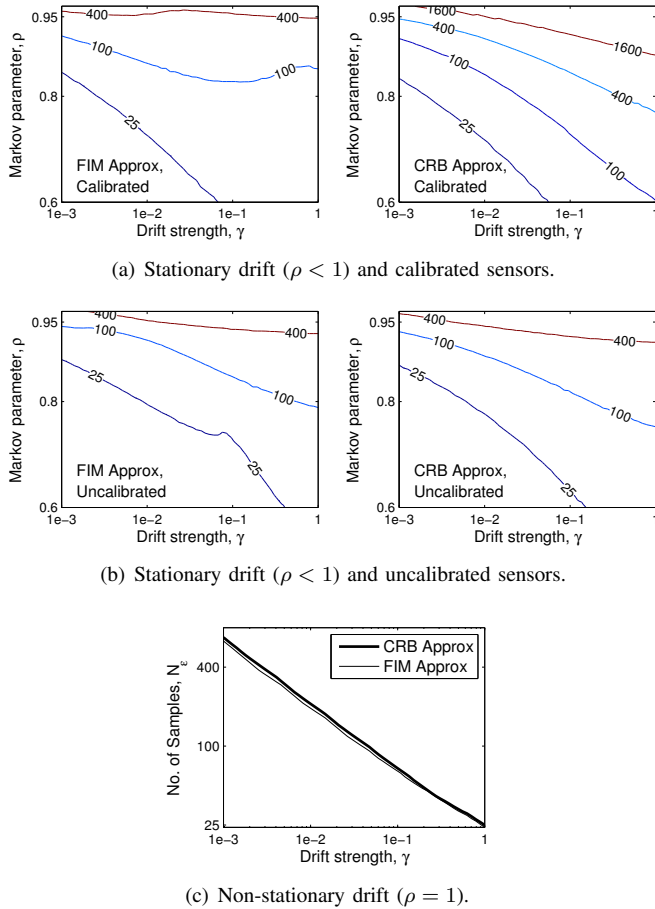
(a) Stationary drift ($\rho < 1$) and calibrated sensors.

(b) Stationary drift ($\rho < 1$) and uncalibrated sensors.

(c) Non-stationary drift ($\rho = 1$).

Fig. 1.   Sample size N_ϵ required for 95% accurate performance prediction

inversely proportional to ρ and γ. For the random-walk drift scenario (Figure 1(c)), higher value of drift-strength γ requires smaller observation durations (N_ϵ) for accurate performance predictions. This too, can be explained from approximation-region condition in (44), where, with the aid of definitions (37) and (39), it can be established that $\epsilon_p = \mathcal{O}(\gamma^{-1/2})$ for small γ. This also partially explains the log-linear relation between N_ϵ and γ in Figure 1(c).

We have so far only considered the estimation performance of a single sensor that is able to reliably communicate its observations to the sink without any distortion. In Sections IV and V, we discuss extensions to the multiple sensor framework and bandwidth limited networks.

IV. MULTIPLE SENSORS

In this section, we consider the application of the results in Section III to multiple sensors with different noise and drift parameters $\{\sigma_m^2, \gamma_m, \rho_m\}$. Since the sensor noise and drift are independent across the sensors (Assumption 3), the Fisher Information for $\boldsymbol{\beta}$ is equal to the sum of individual FIM-s (see equation (15)), $\boldsymbol{J} = \boldsymbol{J}_1 + \boldsymbol{J}_2 + \cdots + \boldsymbol{J}_M$. Hence the expressions for FIM (given by (28) and (38)) and subsequently CRB (given by (33) and (40)) can be extended with the following change

in definitions,

$$\xi_{\#,\text{eff}} \triangleq \sum_{m=1}^{M} \xi_{\#}(\sigma_m^2, \gamma_m, \rho_m), \text{ for } [\#] = 0, 1, \ldots, 6, \quad (46)$$

where $\xi_{\#}(\sigma^2, \gamma, \rho)$ can be thought of as a function of its arguments as defined in (29) and (39) and $\xi_{\#,\text{eff}}$ denote the effective value of the constant.

As an illustration of this extension, we consider an example where noise and drift parameters are uniformly (randomly) distributed over a given range, say,

$$\rho_m \in [\rho_l, \rho_u], \ \sigma_m^2 \in [\sigma_l^2, \sigma_u^2] \text{ and } \gamma_m \in [\gamma_l, \gamma_u]. \quad (47)$$

When the number of sensors M is large, (46) can be further approximated by substituting the summations by integrals,

$$\xi_{\#,\text{eff}} = M \mathbb{E}\left[\xi_{\#}\right]$$
$$= \frac{M}{\Delta\rho\Delta\sigma^2\Delta\gamma} \int_{\sigma_l^2}^{\sigma_u^2} \int_{\gamma_l}^{\gamma_u} \int_{\rho_l}^{\rho_u} \xi_{\#} \, \mathrm{d}\rho \, \mathrm{d}\gamma \, \mathrm{d}\sigma^2. \quad (48)$$

We would refer to the CRB derived using the constants $\xi_{\#,\text{eff}}$ (arising out of integration) as the *Average-CRB*. We expect the Average-CRB to be an effective indicator of system performance for large number of sensors M, as we demonstrate with some numerical results below.
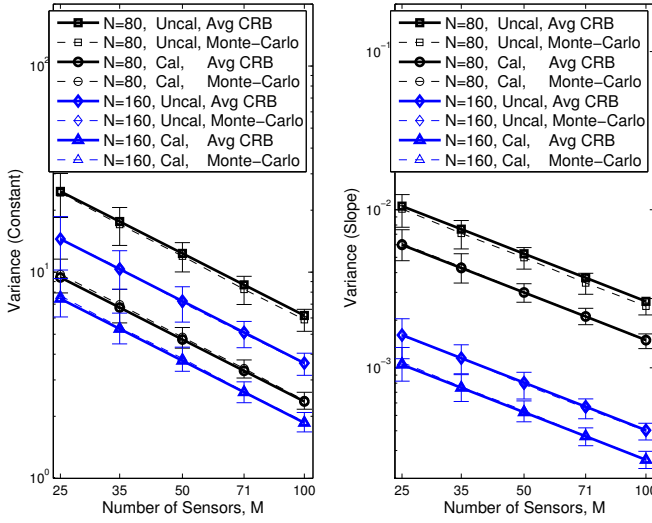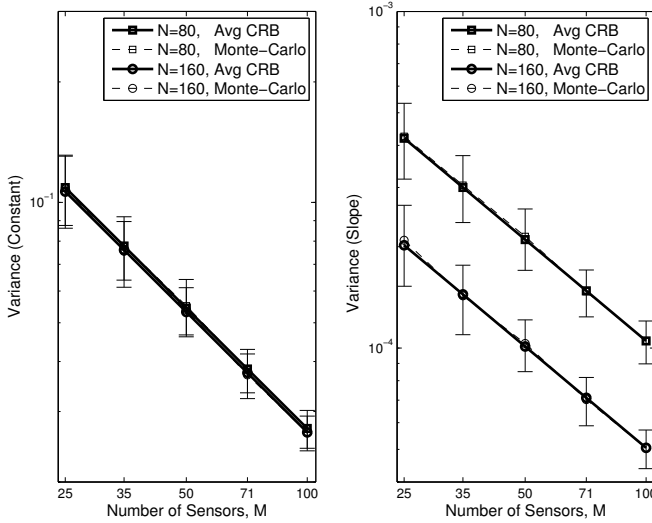
A. Numerical Results

We demonstrate the results for both $\rho < 1$ and $\rho = 1$ cases. The simulation setup is described next. *1)* We consider a linear signal, i.e., $P = 1$. The parameters to be estimated are the constant term β_0 and the slope term β_1. *2)* We consider multi-sensor systems with the number of sensors, M, starting from 25 and going upto 100. *3)* For the $\rho < 1$ scenario, for each sensor, the drift parameters are randomly selected by choosing the parameters from the range,

$$\rho_m \in [.85, .95], \ \sigma_m^2 \in [72, 288] \text{ and } \gamma_m \in [.6, 2.4], \quad (49)$$

which is close to an estimated spectrum in [17]. Within the $\rho < 1$ scenario, we simulate both calibrated ($\tau_m = 1, \forall m$) and uncalibrated ($\tau_m = \infty, \forall m$) cases. In one simulation, we assume that all M sensors are calibrated while on another simulation we assume that none of them are calibrated. *4)* For the $\rho = 1$ scenario, the drift parameters are randomly selected by choosing the parameters from the range,

$$\sigma_m^2 \in [2, 12] \text{ and } \gamma_m \in [.05, .3]. \quad (50)$$

5) For each realization of an M-sensor network, ML-estimation of the constant and slope parameters were performed and the error variances were averaged over 10^5 Monte-Carlo trials (realizing the measurement noise and sensor drift). *6)* Several (10^3) Monte-carlo trials (realizing M-sensor networks with $\sigma^2, \rho_m, \gamma_m$ chosen from above parameter range) are performed and the average and 95% confidence interval of the error variances are observed. *7)* We repeat the experiments for sample sizes $N = 80$ and 160. These samples sizes were chosen to ensure moderate computational (Monte-Carlo) effort. Since we consider a small number of samples, we have used the expressions for Fisher Information Matrix (Equations

Fig. 2.   Average performance for multi-sensor systems for $\rho \in [0.85, 0.95]$.

Fig. 4.   Bandwidth constrained sensor network where inference is performed using quantized data.

Fig. 3.   Average performance for multi-sensor systems for $\rho = 1$.

(28) and (38)) to predict the estimation error variance of the constant β_0 and slope β_1 portion of a signal. The results are displayed in Figures 2 and 3, of which we make some comments below.

Firstly, an approximation to the system performance using (48) is seen to be fairly accurate, as depicted in Figures 2 and 3. In all cases, the variance is inversely proportional to the number of sensors, as depicted by the log-linearity of all the curves. The (average) performance prediction improves (error-bars becomes shorter) for higher sample sizes, since the summation (46) is better approximated by the integral (48) for large M. This means that just by knowing the range of the parameter values and the number of sensors, we can have a good understanding about the estimation performance of the entire sensor network. Secondly, from Figure 2, we note the intuitive phenomenon that the performance using calibrated sensors is better than that using uncalibrated sensors. Also,
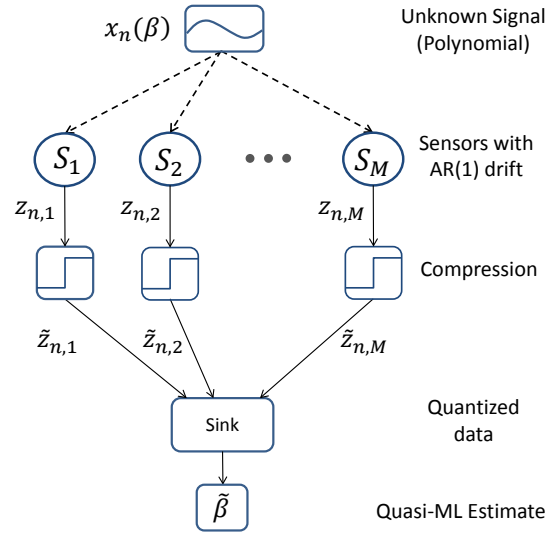
though for $N = 80$, the performance gap is quite large, the gap between calibrated and uncalibrated cases narrows down for $N = 160$. This corroborates Remark 3.4, where we noted that the relative performance gain diminishes for higher sample sizes. For the $\rho = 1$ case, from Figure 3, we note that increasing the sample size does not help in estimation of the constant portion of the signal. This is due the inconsistency property of the non-stationary drift model, as described in Remark 4.1.

V. QUANTIZED OBSERVATIONS

As another application of the results in Section III, we consider the performance characterization of a sensor network where resource (bandwidth, power, etc.) available for communication is limited [2], [32]. Bandwidth constraints preclude the reliable communication of full-precision observations from the individual sensor nodes to the sink. In such situations, the observations must be compressed and digitized prior to transmission [38]. In this section, we digitize each observation by using uniform quantization, as will be discussed shortly. Since the reliability of the transmission of digitized observations may be further enhanced through the use of efficient channel coding (error-control codes), it is reasonable to assume that once the quantization is performed, there is no further deterioration of the observation-quality [2]. In other words, the quantized observations are assumed to be available error-free at the sink. The schematic diagram of such a system is depicted in Figure 4. *Our goal in this section is to predict the estimation performance of the sensor network composed of sensors with drift, when only quantized observations are available for inference at the sink.*

We use uniform quantization [33], in which each of the observations $z_{n,m}$ are quantized uniformly with l-bits. Assuming that the observations are bounded $z_{n,m} \in [U_0, U_1]$ and the quantization thresholds $a_j \in [U_0, U_1]$, $j = 0, \ldots, 2^l$ are uniformly spaced such that $a_{j+1} - a_j = (U_1 - U_0)/(2^l - 1) \triangleq \Delta$,

the expected distortion due to the quantization process is

$$\sigma_Q^2 \triangleq \mathbb{E}[(\widetilde{z}_{n,m} - z_{n,m})^2] = \frac{\Delta^2}{12}, \quad \forall n, m. \qquad (51)$$

Since the quantization noise $\widetilde{z}_{n,m} - z_{n,m}$, in general, is neither independent across time nor Gaussian distributed, the Maximum-Likelihood estimator is difficult to design. However, the quasi-ML estimator is easier to implement, which is designed on the assumption that the quantization noise is i.i.d. Gaussian with variance σ_Q^2. We will use the quasi-ML estimator, described below, to perform inference using quantized observations.

Effectively, the quasi-ML estimator assumes that the noise at individual sensor nodes has an added component $\sigma_Q^2 \boldsymbol{I}$, with the total covariance being (compare with (13))

$$\begin{aligned}
\widetilde{\boldsymbol{\Sigma}}_{*,m} &\triangleq \sigma_Q^2 \boldsymbol{I} + \boldsymbol{\Sigma}_{*,m} \\
&= \sigma_Q^2 \boldsymbol{I} + \sigma_m^2 (\boldsymbol{I} + \gamma_m \boldsymbol{R}_{*,m}) \\
&\triangleq \widetilde{\sigma}_m^2 (\boldsymbol{I} + \widetilde{\gamma}_m \boldsymbol{R}_{*,m}),
\end{aligned} \qquad (52)$$

where $\widetilde{\sigma}_m, \widetilde{\gamma}_m$ satisfy $\widetilde{\sigma}_m^2 \triangleq \sigma_m^2 + \sigma_Q^2$ and $\widetilde{\gamma}_m \triangleq \frac{\gamma_m}{1 + \sigma_Q^2/\sigma_m^2}$. Accordingly, the quasi-ML estimator is (compare to (14)),

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}}^{\text{ML}} &= \widetilde{J}^{-1} \boldsymbol{X}' \sum_{m=1}^{M} \widetilde{\boldsymbol{\Sigma}}_m^{-1} \widetilde{\boldsymbol{z}}_m, \text{where} \\
\widetilde{J} &\triangleq \boldsymbol{X}^T \left(\sum_{m=1}^{M} \widetilde{\boldsymbol{\Sigma}}_m^{-1} \right) \boldsymbol{X}.
\end{aligned} \qquad (53)$$

Note that, in the limiting case when quantization errors are small (equivalently, large l), we have $\widetilde{\sigma}_m^2 \to \sigma_m^2, \widetilde{\gamma}_m \to \gamma_m$ and the quasi-ML-estimator is identical to the ML-estimator.

We would refer to the CRB derived using modified noise parameters $\{\widetilde{\sigma}_m^2, \widetilde{\gamma}_m, \rho_m\}$ as the *Modified-CRB*. We expect the Modified-CRB to be an effective indicator of system performance for moderate to large number of quantization levels 2^l. Similar to Section IV, with certain change in definitions, the expressions for FIM (given by (28) and (38)) and subsequently CRB (given by (33) and (40)) can be extended to obtain the Modified-CRB, $\xi_{\#,\text{eff}} \triangleq \sum_{m=1}^{M} \xi_{\#}(\rho_m, \widetilde{\sigma}_m^2, \widetilde{\gamma}_m)$, where $\xi_{\#}(\rho, \sigma^2, \gamma)$ can be thought of as a function of its arguments as defined in (29) and (39) for $[\#] = 0, 1, \ldots, 6$. We show some numerical results below to corroborate the effectiveness of Modified-CRB as an efficient performance predictor.

A. Numerical Results

The simulation setup is described as follows. *1)* A sample size of $N = 400$ and network size of $M = 25$ nodes was considered. *2)* The noise and drift parameters of the sensor nodes are chosen by uniformly spacing them in the range,

$$72 = \sigma_1^2 < \sigma_2^2 < \cdots < \sigma_M^2 = 288,$$
$$0.6 = \gamma_1 < \gamma_2 < \cdots < \gamma_M = 2.4, \text{and}$$
$$0.85 = \rho_1 < \rho_2 < \cdots < \rho_M = 0.95.$$

Both calibrated ($\tau_m = 1, \forall m$) and uncalibrated ($\tau_m = \infty, \forall m$) cases were considered. *3)* We assume a linear signal, i.e., $P = 1$ with the constant term $\beta_0 = 400$ and the slope term $\beta_1 = 0.9$. The range of the observations
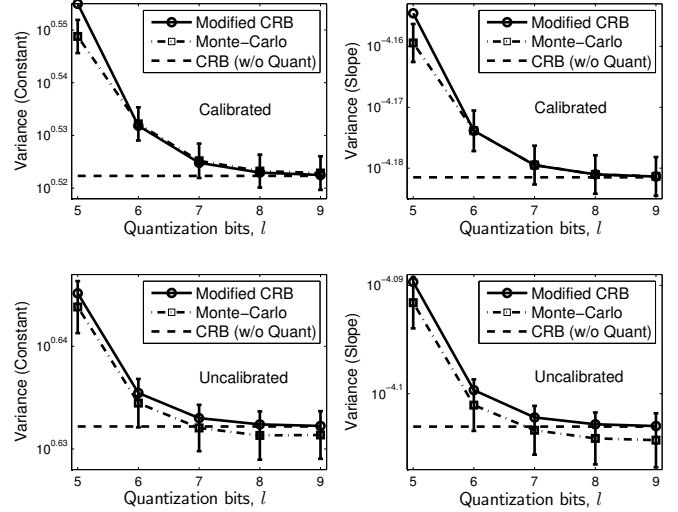

Fig. 5. Cramér-Rao bounds with quantized observations.

to be quantized was chosen to be $U_0 = 0, U_1 = 1200$, beyond which the observations were clipped. The range was deliberately chosen large so that clipping (which is another source of distortion which we have not modeled) does not occur frequently. *4)* Uniform quantization is performed using $l = 5, 6, \ldots, 9$ bits per observation. *5)* Quasi-ML estimation of the constant and slope parameters was performed and the error variances were averaged over 10^8 Monte-Carlo trials (realizing the measurement noise and sensor drift). The 95% confidence interval of the error variances were also observed.

The results of Monte-Carlo simulations are compared to theoretical predictions from Modified-CRB and displayed in Figure 5. The full-precision CRB is also displayed in the figure (labeled w/o Quant), marking the convergence of Modified-CRB in the large-l regime. The actual estimation variance of both the constant and slope parameters (of the linear signal) seem to agree, with reasonable accuracy, to the Modified-CRB.

VI. CONCLUSION

In this paper, we have derived approximate bounds for the estimation accuracy of polynomial signals using sensors that exhibit drift in addition to having measurement errors. This is important since sensor drift, or loss of calibration with time, is a major problem in many applications. The theoretical closed form expressions are validated through numerical results. As future work, we intend to derive performance measures for other signal models, e.g., stochastic models rather than deterministic. Additionally, it may be interesting to investigate the impact of spatial correlation (i.e, relaxing Assumption 3) on estimation performance. More bit-efficient quantization schemes (other than uniform quantization) that optimally allocate bandwidth to each of the sensors is also another topic of interest. Finally, in-network inference where only the summary of estimated parameters are communicated to the sink, rather than entire observations, will be another framework that one might consider.

APPENDIX A
PROOF OF PROPOSITION 1

In Proposition 1, we need to show equation (17), which is $(I + \gamma R)^{-1} = I - \nu M + \mathcal{O}(y^N)$. Define $\widetilde{M}$ be such that $(I + \gamma R)^{-1} = I - \nu \widetilde{M}$, i.e., the exact form of (17). It suffices to prove that $M = \widetilde{M} + \mathcal{O}(y^N)$, or equivalently

$$\widetilde{M}^{-1} M = I + \mathcal{O}(y^N). \tag{54}$$

To prove (54), we need $\widetilde{M}^{-1}$. Definition of $\widetilde{M}$ implies

$$\widetilde{M}^{-1} = \nu(I + R^{-1}/\gamma). \tag{55}$$

The inverse of R_U (given by (12)) is well known [34]. The inverse of R is similar to that of R_U except for the top-left element. Specifically,

$$R^{-1} = \begin{bmatrix} 1 + \varrho_\tau & -\rho & \cdots & 0 & 0 \\ -\rho & 1 + \rho^2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 + \rho^2 & -\rho \\ 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}. \tag{56}$$

From (55) and with the help of some identities that follow from (19), namely $\nu(1 + 1/\gamma) = (1 - \rho y + y^2)/(1 - y^2)$ and $\nu(1 + (1 + \rho^2)/\gamma) = (1 + y^2)/(1 - y^2)$, we have,

$$\widetilde{M}^{-1} = \frac{1}{1 - y^2} \begin{bmatrix} m_\tau & -y & \cdots & 0 & 0 \\ -y & 1 + y^2 & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 + y^2 & -y \\ 0 & 0 & \cdots & -y & 1 - \rho y + y^2 \end{bmatrix},$$

where $m_\tau \triangleq 1 - \rho y + y^2 + \varrho_\tau y/\rho$. Constructing M from Equation (21), we can now directly verify (54). Specifically, the residual is of the following structure

$$\widetilde{M}^{-1} M - I = \frac{y^N}{r_0} \begin{bmatrix} 0 & 0 & \cdots & r_1 & r_3 \\ 0 & 0 & \cdot^{\cdot^\cdot} & r_2 & 0 \\ \vdots & \cdot^{\cdot^\cdot} & \cdot^{\cdot^\cdot} & \cdot^{\cdot^\cdot} & \vdots \\ r_1 & r_2 & \cdot^{\cdot^\cdot} & 0 & 0 \\ r_2 & 0 & \cdots & 0 & 0 \end{bmatrix}, \tag{57}$$

where $r_0 \triangleq (1 - \rho y)(\rho - \rho^2 y + \varrho_\tau y)$, $r_1 \triangleq -\varrho_\tau y$, $r_2 \triangleq -\rho(\rho - y)(1 - \rho y) + \varrho_\tau(1 + y^2 - \rho y)$ and $r_3 \triangleq r_2 + \varrho_\tau y(\varrho_\tau - \rho^2)/\rho$. Since (57) are terms of order $\mathcal{O}(y^N)$, this completes the proof of (54) and hence Proposition 1.

APPENDIX B
POLYNOMIAL APPROXIMATION OF $X'MX$

First, we would decompose M (see (21)) to aid analytical calculations. Note that

$$M = I + M_1 + \eta_\tau M_2 + \kappa M_3, \tag{58}$$

and matrices M_1, M_2 and M_3 are defined by

$$M_1 \triangleq \begin{bmatrix} 0 & y & \cdots & y^{N-2} & y^{N-1} \\ y & 0 & \ddots & y^{N-3} & y^{N-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ y^{N-2} & y^{N-3} & \ddots & 0 & y \\ y^{N-1} & y^{N-2} & \cdots & y & 0 \end{bmatrix}, \tag{59}$$

$$M_2 \triangleq \begin{bmatrix} 1 & y & \cdots & y^{N-2} & 0 \\ y & \cdot^{\cdot^\cdot} & \cdot^{\cdot^\cdot} & \cdot^{\cdot^\cdot} & \vdots \\ \vdots & \cdot^{\cdot^\cdot} & \cdot^{\cdot^\cdot} & \cdot^{\cdot^\cdot} & \vdots \\ y^{N-2} & \cdot^{\cdot^\cdot} & \cdot^{\cdot^\cdot} & \cdot^{\cdot^\cdot} & 0 \\ 0 & \cdots & \cdots & 0 & 0 \end{bmatrix}, \text{ and} \tag{60}$$

$$M_3 \triangleq \begin{bmatrix} 0 & 0 & \cdots & 0 & y^{N-1} \\ 0 & \cdot^{\cdot^\cdot} & \cdot^{\cdot^\cdot} & y^{N-1} & y^{N-2} \\ \vdots & \cdot^{\cdot^\cdot} & \cdot^{\cdot^\cdot} & \cdot^{\cdot^\cdot} & \vdots \\ 0 & y^{N-1} & \cdot^{\cdot^\cdot} & \cdot^{\cdot^\cdot} & y \\ y^{N-1} & y^{N-2} & \cdots & y & 1 \end{bmatrix}, \tag{61}$$

respectively. To calculate $X'MX$, we compute each of terms in (58). $X'X$ is given by (24) and (25). For $X'M_1 X$, we refer to (59) and collect the identical powers of y from either side of the principal diagonal,

$$[X'M_1 X]_{k,l} = \sum_{i=1}^{N-1} y^i \sum_{j=1}^{N-i} (j^k(i+j)^l + j^l(i+j)^k)$$

$$\overset{(a)}{=} \sum_{i=1}^{N-1} y^i \sum_{j=1}^{N-i} \sum_{r=0}^{k+l} A_r i^r j^{k+l-r}$$

$$= \sum_{i=1}^{N-1} y^i \sum_{r=0}^{k+l} A_r i^r \sum_{j=1}^{N-i} j^{k+l-r}$$

$$\overset{(b)}{=} \sum_{i=1}^{N-1} y^i \sum_{r=0}^{k+l} A_r i^r \sum_{s=0}^{k+l-r} B_{s,r}(N-i)^{k+l-r+1-s}$$

$$\overset{(c)}{=} \sum_{i=1}^{N-1} y^i \sum_{r=0}^{q-1} A_r i^r \sum_{s=0}^{q-r-1} B_{s,r} \sum_{t=0}^{q-r-s} C_{t,s,r} N^{q-r-s-t}(-i)^t$$

$$\overset{(d)}{=} \sum_{i=1}^{N-1} y^i \sum_{r=0}^{q-1} \sum_{s=0}^{q-r-1} \sum_{t=0}^{q-r-s} X_{r,s,t}$$

$$\overset{(e)}{=} \sum_{i=1}^{N-1} y^i \left(\sum_{u=0}^{q-1} \sum_{v=0}^{u} \sum_{r=0}^{v} X_{r,u-v,v-r} + \sum_{v=1}^{q} \sum_{r=0}^{v-1} X_{r,q-v,v-r} \right)$$

$$\overset{(f)}{=} \sum_{i=1}^{N-1} y^i \left(\sum_{u=0}^{q-1} N^{q-u} \sum_{v=0}^{u} i^v \sum_{r=0}^{v} D_{r,v,u} + \sum_{v=1}^{q} i^v \sum_{r=0}^{v-1} D_{r,v,q} \right)$$

$$\overset{(g)}{=} \sum_{u=0}^{q-1} N^{q-u} \underbrace{\sum_{v=0}^{u} Y_v \sum_{r=0}^{v} D_{r,v,u}}_{\triangleq \mathcal{A}_{u,M_1}, \ u \leq k+l} + \underbrace{\sum_{v=1}^{q} Y_v \sum_{r=0}^{v-1} D_{r,v,q}}_{\triangleq \alpha_{M_1}} + \mathcal{O}(N^q y^N),$$

$$\tag{62}$$

where in step (a) we collect all like powers of i and j and define $A_r \triangleq \binom{l}{r} + \binom{k}{r}$, (b) follows from the summation formula in (25) and the definition $B_{s,r} \triangleq \mathcal{B}_{k+l-r,s}$, (c) is due to binomial expansion and definitions $q \triangleq k + l + 1$ and $C_{t,s,r} \triangleq \binom{q-r-s}{t}$. Step (d) follows from the definition $X_{r,s,t} \triangleq i^{r+t} N^{q-r-s-t} A_r B_{s,r} C_{t,s,r} (-1)^t$, step (e) is an identity involving series rearrangement (i.e., true for any $X_{r,s,t}$, by defining $u \triangleq r+s+t$ and $v \triangleq r+t$), step (f) follows from defining $D_{r,v,u} \triangleq A_r B_{u-v,r} C_{v-r,u-v,r} (-1)^{v-r}$. Step (g) follows from the definition $Y_v \triangleq \sum_{i=1}^{\infty} i^v y^i$ (which converges for finite v since $y < 1$ and the error term is of the order $N^v y^N$). Quantities of the form Y_v are also known as poly-logarithms. In the last step, we denote the polynomial coefficients by $\mathcal{A}_{u,M_1}$ and the constant term by α_{M_1}.

For $\boldsymbol{X}'\boldsymbol{M}_2\boldsymbol{X}$, we start with (60) and collect the identical powers of y from the top-left half,

$$[\boldsymbol{X}'\boldsymbol{M}_2\boldsymbol{X}]_{k,l} = \sum_{i=1}^{N-1} y^{i-1} \sum_{j=1}^{i} j^k (i+1-j)^l$$
$$= \underbrace{\sum_{i=1}^{\infty} y^{i-1} \sum_{j=1}^{i} j^k (i+1-j)^l}_{\triangleq \alpha_{M_2}} + \mathcal{O}(N^q y^N), \quad (63)$$

where α_{M_2} is a constant. Similarly, for $\boldsymbol{X}'\boldsymbol{M}_3\boldsymbol{X}$, we collect the identical powers of y from the bottom-right half of (61),

$$[\boldsymbol{X}'\boldsymbol{M}_3\boldsymbol{X}]_{k,l} = \sum_{i=1}^{N-1} y^{i-1} \sum_{j=1}^{i} (N-j+1)^k (N+j-i)^l$$
$$\stackrel{(a)}{=} \sum_{i=1}^{N-1} y^{i-1} \sum_{j=1}^{i} \sum_{r=0}^{k} G_r N^{k-r} (-j+1)^r \sum_{s=0}^{l} H_s N^{l-s} (j-i)^s$$
$$\stackrel{(b)}{=} \sum_{i=1}^{N-1} y^{i-1} \sum_{j=1}^{i} \sum_{t=0}^{k+l} N^{k+l-t} \sum_{s=0}^{t} K_{t,s}$$
$$\stackrel{(c)}{=} \sum_{t=0}^{k+l} N^{k+l-t} \mathcal{A}_{t,M_3} + \mathcal{O}(N^q y^N), \quad (64)$$

where (a) follows from binomial expansion and definitions $G_r \triangleq \binom{k}{r}$, $H_s \triangleq \binom{l}{s}$. In (b), we have rearranged the sum to group similar exponents of N with the transformation $t \triangleq r+s$ and defined $K_{t,s} \triangleq G_{t-s} (-j+1)^{t-s} H_s (j-i)^s$, (c) follows from similar arguments while deriving (62) and (63) and the definition $\mathcal{A}_{t,M_3} \triangleq \sum_{i=1}^{\infty} y^{i-1} \sum_{j=1}^{i} \sum_{s=0}^{t} K_{t,s}$, $t \leq k+l$.

We note that all the constants $\mathcal{A}_{u,M_1}, \alpha_{M_1}, \alpha_{M_2}, \mathcal{A}_{t,M_3}$ depend on k and l. We can now prove Proposition 2 by composing $\boldsymbol{X}'\boldsymbol{M}\boldsymbol{X}$ with the help of (62), (63) and (64) in (58). Define $\mathcal{A}_{-1,M_3} = 0$. The constants $\mathcal{A}_{k,l,i}$ and $\alpha_{k,l}^{(\tau)}$ in Proposition 2 are given by

$$\mathcal{A}_{k,l,i} = \mathcal{B}_{k+l,i} + \mathcal{A}_{i,M_1} + \kappa \mathcal{A}_{i-1,M_3}, \quad 0 \leq i \leq k+l,$$
$$\alpha_{k,l}^{(\tau)} = \alpha_{M_1} + \eta_\tau \alpha_{M_2} + \kappa \mathcal{A}_{k+l,M_3}, \quad (65)$$

respectively. Some of these constants are enumerated in Table I. This completes the proof. For example, to compute $\mathcal{A}_{k,l,0}$, note that $\mathcal{B}_{k+l,0} = \frac{1}{k+l+1}$, $\mathcal{A}_{0,M_1} = \frac{2Y_0}{k+l+1}$, $Y_0 = \frac{y}{1-y}$, hence

$$\mathcal{A}_{k,l,0} = \mathcal{B}_{k+l,0} + Y_0 \mathcal{A}_{0,M_1} = \frac{1+y}{1-y} \frac{1}{k+l+1}, \quad (66)$$

which agrees with (27).

APPENDIX C
FIM AND CRB FOR $\rho = 1$

First we derive the FIM given by Equation (38). Let the Taylor Series of the FIM be of the form

$$\boldsymbol{J} = \frac{1}{\sigma^2} \boldsymbol{E} \left[\boldsymbol{Q}_0 + \frac{1}{N} \boldsymbol{Q}_1 + \frac{1}{N^2} \boldsymbol{Q}_2 + \mathcal{O}\left(\frac{1}{N^3}\right) \right] \boldsymbol{E}. \quad (67)$$

The first term $\boldsymbol{Q}_0$ is equal to $\xi_2 \boldsymbol{f}\boldsymbol{f}'$ (follows from (28) and the fact that $\xi_0 = \xi_1 = 0$ for $\rho = 1$). The second order term $\boldsymbol{Q}_1$ follows from collecting third-order terms in (24) and (26) and adding them according to (23), i.e.,

$$[\boldsymbol{Q}_1]_{k,l} = \begin{cases} 0, & k = l = 0, \\ -\nu \alpha_{k,l}^{(\tau)}, & k+l = 1, \\ \mathcal{B}_{k+l,2} - \nu \mathcal{A}_{k,l,2}, & k+l \geq 2. \end{cases} \quad (68)$$

From definitions (19) and identity (22), we obtain the following simplifications for $\rho = 1$, $\nu = \frac{1-y}{1+y}$, $\kappa = y$ and $\frac{1}{\gamma} = \frac{y}{(1-y)^2}$, which when applied to (65) (and using values from Table I) leads to $[\boldsymbol{Q}_1]_{k,l} = \frac{kl}{k+l-1} \frac{1}{\gamma}$ for $k + l \geq 2$, which explains the term ξ_4 in (39). Similarly, $\boldsymbol{Q}_2$ follows from collecting fourth-order terms,

$$[\boldsymbol{Q}_2]_{k,l} = \begin{cases} 0, & k+l \leq 1, \\ -\nu \alpha_{k,l}^{(\tau)}, & k+l = 2, \\ \mathcal{B}_{k+l,3} - \nu \mathcal{A}_{k,l,3}, & k+l \geq 3, \end{cases} \quad (69)$$

the last term of which when simplified yields $[\boldsymbol{Q}_2]_{k,l} = -kl \frac{y^2}{(1-y)^3}$ for $k + l \geq 3$, that explains the term ξ_5 in (39). The derivation of the CRLB in Equation (40) follows from the block inversion of $\boldsymbol{J}$ and subsequent application of Lemma 1,

$$\begin{bmatrix} a & \boldsymbol{b}' \\ \boldsymbol{b} & \boldsymbol{C} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{a} + \frac{\boldsymbol{b}'[\boldsymbol{C}-\frac{1}{a}\boldsymbol{b}\boldsymbol{b}']^{-1}\boldsymbol{b}}{a^2} & * \\ * & [\boldsymbol{C} - \frac{1}{a}\boldsymbol{b}\boldsymbol{b}']^{-1} \end{bmatrix}$$
$$\triangleq \begin{bmatrix} m & * \\ * & \boldsymbol{M} \end{bmatrix}, \text{ say.} \quad (70)$$

From (38), we have $a = \xi_2^{(\tau)}$, $\boldsymbol{b} = \frac{1}{N}\xi_3^{(\tau)} \boldsymbol{f}' + \mathcal{O}\left(\frac{1}{N^2}\right)$ and

$$\boldsymbol{C} = \frac{1}{N}\boldsymbol{D} \left[\xi_4 \boldsymbol{H} + \frac{1}{N}\left(\xi_5 \boldsymbol{e}\boldsymbol{e}' + \xi_6^{(\tau)} \boldsymbol{f}\boldsymbol{f}' \right) + \mathcal{O}\left(\frac{1}{N^2}\right) \right] \boldsymbol{D},$$

where subscript P is dropped from $\boldsymbol{H}_P, \boldsymbol{e}_P$ and $\boldsymbol{f}_P$. Next we apply Lemma 1 to obtain the diagonal elements $[\boldsymbol{M}]_{q,q}$ of block $\boldsymbol{M}$. By substituting $c_0 = \xi_4$, $c_1 = \xi_5$ $c_2 = \xi_6^{(\tau)} - \left(\xi_3^{(\tau)}\right)^2/\xi_2^{(\tau)}$, and observing the facts that $[\boldsymbol{D}^{-1}]_{p,p} = \frac{1}{p}$ and $L_{P-1,p-1} = P^2/p^2$, we obtain $[\boldsymbol{V}]_{p,p}$ for $p \geq 1$ as in (40). For the top-left element, m, the second order term is

$$\frac{\boldsymbol{b}'\boldsymbol{M}\boldsymbol{b}}{a^2} = \frac{\left(\xi_3^{(\tau)}\right)^2}{\left(\xi_2^{(\tau)}\right)^2} \boldsymbol{M}_{1,1}, \text{ but } \boldsymbol{M}_{1,1} = \frac{K_{P-1,0}}{N\xi_4} + \mathcal{O}\left(\frac{1}{N^2}\right),$$

which completes the derivation of (40).

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *Communications Magazine, IEEE*, vol. 40, no. 8, pp. 102–114, 2002.

[2] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-Part I: Gaussian case," *Signal Processing, IEEE Transactions on*, vol. 54, no. 3, pp. 1131–1143, 2006.

$i\downarrow$	$\mathcal{B}_{k+l,i}$	$\mathcal{A}_{i,M_1}$	$\mathcal{A}_{i-1,M_3}$	Y_i
0	$\frac{1}{k+l+1}$	$\frac{2}{k+l+1}Y_0$	0	$\frac{y}{1-y}$
1	$\frac{1}{2}$	$Y_0 - Y_1$	$\frac{1}{(1-y)^2}$	$\frac{y}{(1-y)^2}$
2	$\frac{k+l}{12}$	$\frac{k+l}{6}Y_0 - \frac{k+l}{2}Y_1 + \frac{\binom{k}{2}+\binom{l}{2}}{k+l-1}Y_2$	$-\frac{(k+l)y}{(1-y)^3}$	$\frac{y(1+y)}{(1-y)^3}$
3	0	$-\frac{\binom{k+l}{2}}{6}Y_1 + \frac{\binom{k}{2}+\binom{l}{2}}{2}Y_2 - \frac{\binom{k}{2}+\binom{l}{2}-\frac{kl}{2}}{3}Y_3$	$\frac{y\left((1+y)\left(\binom{k}{2}+\binom{l}{2}\right)+kly\right)}{(1-y)^4}$	$\frac{y(1+4y+y^2)}{(1-y)^4}$

$(k,l)\downarrow$	α_{M_1}	α_{M_2}	$\mathcal{A}_{k+l,M_3}$
$(0,0)$	$-\frac{2y}{(1-y)^2}$	$\frac{1}{(1-y)^2}$	$\frac{1}{(1-y)^2}$
$(0,1),(1,0)$	$-\frac{y}{(1-y)^2}$	$\frac{1}{(1-y)^3}$	$-\frac{y}{(1-y)^3}$
$(1,1)$	$\frac{2y^2}{(1-y)^4}$	$\frac{1}{(1-y)^4}$	$\frac{y^2}{(1-y)^4}$

TABLE I

VALUES OF SEVERAL CONSTANTS HELPFUL TO DERIVE $\mathcal{A}_{k,l,i}$ AND $\alpha_{k,l}^{(\tau)}$.

[3] R. Niu and P. Varshney, "Target location estimation in sensor networks with quantized data," *Signal Processing, IEEE Transactions on*, vol. 54, no. 12, pp. 4519–4528, Dec. 2006.

[4] Y. Bar-Shalom, "Airborne GMTI radar position bias estimation using static-rotator targets of opportunity," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 37, no. 2, pp. 695–699, Apr. 2001.

[5] G. A. Washer, P. A. Fuchs, M. K. Masterson, and C. Phillips, "Long-term structural health monitoring for bridge piers," *ASNT Research Symposium and Spring Conference, Anaheim, CA*, Mar 31–Apr 3, 2008.

[6] W. J. Fisk, "A pilot study of the accuracy of CO_2 sensors in commercial buildings," *Proceedings of IAQ, Healthy and Sustainable Buildings, Baltimore, Maryland*, 2007.

[7] E. Oka, "Long-term sensor drift found in recovered argo profiling floats," *Journal of Oceanography*, vol. 61, no. 4, pp. 775–781, 2005.

[8] M. Holmberg and T. Artursson, "Drift compensation, standards, and calibration methods," in *Handbook of Machine Olfaction: Electronic Nose Technology*, T. C. Pearce, S. S. Schiffman, H. T. Nagle, and J. W. Gardner, Eds. Weinheim, FRG: Wiley, 2004, ch. 13, pp. 325–346.

[9] M. Takruri, S. Rajasegarar, S. Challa, C. Leckie, and M. Palaniswami, "Online drift correction in wireless sensor networks using spatio-temporal modeling," *Proc. 11th Intl. Conf. Information Fusion, Cologne, Germany*, pp. 158–165, 2008.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[11] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. Wiley-Interscience, 2001.

[12] G. Wei, "Estimations of frequency and its drift rate," *Instrumentation and Measurement, IEEE Transactions on*, vol. 46, no. 1, pp. 79–82, Feb. 1997.

[13] T. C. Pearce and J. W. Gardner, "Predicting organoleptic scores of sub-ppm flavour notes: Part 2. computational analysis and results," *Analyst*, vol. 123, pp. 2057–2066, 1998.

[14] W. Tsujita, H. Ishida, and T. Moriizumi, "Dynamic gas sensor network for air pollution monitoring and its auto-calibration," in *Sensors, 2004. Proceedings of IEEE*, Oct. 2004, pp. 56–59.

[15] K. Kastella, B. Yeary, T. Zadra, R. Brouillard, and E. Frangione, "Bias modelling and estimation for GMTI applications," *Proc. 3rd Intl. Conf. Information Fusion, Paris, France*, 2000.

[16] S. M. Seong, J. G. Lee, and C. G. Park, "Equivalent ARMA model representation for RLG random errors," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 36, no. 1, pp. 286–290, Jan. 2000.

[17] Y. Hayashi and R. Matsuda, "Deductive prediction of measurement precision from signal and noise in liquid chromatography," *Analytical Chemistry*, vol. 66, no. 18, pp. 2874–2881, 1994.

[18] C. G. Jakobson, M. Feinsod, and Y. Nemirovsky, "Low frequency noise and drift in Ion Sensitive Field Effect Transistors," *Sensors and Actuators B: Chemical*, vol. 68, no. 1-3, pp. 134–139, 2000.

[19] G. Hazel, F. Bucholtz, and I. D. Aggarwal, "Characterization and modeling of drift noise in fourier transform spectroscopy: Implications for signal processing and detection limits," *Applied Optics*, vol. 36, no. 27, pp. 6751–6759, 1997.

[20] M. Unser, "Splines: A perfect fit for signal and image processing," *Signal Processing Magazine, IEEE*, vol. 16, no. 6, pp. 22–38, Nov. 1999.

[21] M. Vetterli, P. Marziliano, and T. Blu, "A sampling theorem for periodic piecewise polynomial signals," in *IEEE Proc. Intl. Conf. on Acoustics, Speech, and Signal Processing, 2001. (ICASSP '01)*, 2001, pp. 3893–3896.

[22] G. Wang, J. Cao, H. Wang, and M. Guo, "Polynomial regression for data gathering in environmental monitoring applications," in *Global Telecommunications Conf., IEEE Globecom*, 2007, pp. 1307–1311.

[23] Y. V. Zakharov, T. C. Tozer, and J. F. Adlard, "Polynomial spline-approximation of Clarke's model," *Signal Processing, IEEE Transactions on*, vol. 52, no. 5, pp. 1198–1208, May 2004.

[24] M. P. Dana, "Registration: A prerequisite for multiple sensor tracking," in *Multitarget-Multisensor Tracking: Advanced Applications*, Y. Bar-Shalom, Ed. Norwood, MA: Artech House, 1990, ch. 5, pp. 155–185.

[25] Y. Zhou, H. Leung, and P. C. Yip, "An exact maximum likelihood registration algorithm for data fusion," *Signal Processing, IEEE Transactions on*, vol. 45, no. 6, pp. 1560–1573, June 1997.

[26] D. P. Dee and R. Todling, "Data assimilation in the presence of forecast bias: The GEOS moisture analysis," *Monthly Weather Review*, vol. 128, no. 9, pp. 3268–3282, Sept. 2000.

[27] D. Zupanski and M. Zupanski, "Model error estimation employing an ensemble data assimilation approach," *Monthly Weather Review*, vol. 134, no. 5, pp. 1337–1354, May 2006.

[28] M. Takruri, S. Rajasegarar, S. Challa, C. Leckie, and M. Palaniswami, "Spatio-temporal modelling-based drift-aware wireless sensor networks," *Wireless Sensor Systems, IET*, vol. 1, no. 2, pp. 110–122, June 2011.

[29] T. Li, H. Zhou, and O. M. Collins, "Estimating the polynomial channel," *International Symposium on Information Theory and its Applications, Parma, Italy*, 2004.

[30] S. Peleg and B. Porat, "The Cramer-Rao lower bound for signals with constant amplitude and polynomial phase," *Signal Processing, IEEE Transactions on*, vol. 39, no. 3, pp. 749–752, Mar 1991.

[31] A. Swami, "Cramer-Rao bounds for deterministic signals in additive and multiplicative noise," *Signal Processing*, vol. 53, no. 2-3, pp. 231–244, 1996.

[32] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *Information Theory, IEEE Transactions on*, vol. 51, no. 6, pp. 2210–2219, June 2005.

[33] A. Krasnopeev, J.-J. Xiao, and Z.-Q. Luo, "Minimum energy decentralized estimation in a wireless sensor network with correlated sensor noises," *EURASIP J. on Wireless Communications and Networking*, vol. 4, pp. 473–482, 2005.

[34] R. Niu and P. K. Varshney, "Sampling schemes for sequential detection with dependent observations," *Signal Processing, IEEE Transactions on*, vol. 58, no. 3, pp. 1469–1481, Mar. 2010.

[35] L. J. Skidmore, "Discrete and continous estimation in correlated noise with finite observation time." *Dissertation (Ph.D.), California Institute of Technology.*, 1964.

[36] R. F. Galbraith and J. I. Galbraith, "On the inverses of some patterned matrices arising in the theory of stationary time series," *Journal of Applied Probability*, vol. 11, no. 1, pp. 63–71, 1974.

[37] I. S. Gradshtein, I. M. Ryzhik, A. Jeffrey, and D. Zwillinger, *Table of Integrals, Series and Products*. Amsterdam: Elsevier, 2007.

[38] A. Goldsmith, *Wireless Communications*. Cambridge: Cambridge University Press, 2005.

Swarnendu Kar (S'07) received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, India, in 2004 and the M.S. degree in mathematics from Syracuse University, Syracuse, NY, in 2009, where he is currently working toward the Ph.D. degree in electrical engineering. He was a Video Systems Engineer with Ittiam Systems Pvt. Ltd., Bangalore, India, during 2004–2006. He was a visiting student at The University of Melbourne, Parkville, Australia, during October–December 2009. His research interests include detection and estimation theory and distributed estimation in the context of sensor networks.

Marimuthu Palaniswami received his ME from the Indian Institute of Science, India, MEngSc from the University of Melbourne and PhD from the University of Newcastle, Australia before rejoining the University of Melbourne. He has published over 350 refereed research papers in highly reputed journals and conferences. He currently leads a large ARC Research Network on Intelligent Sensors, Sensor Networks, and Information Processing (ISSNIP), which is a network centre of excellence with complementary funding for fundamental research, test beds, international linkages, and industry linkages. Previously, he was a director of Centre of Networked Decision Systems (CENDS). His leadership includes external reviewer to an International Research Centre, selection panel member for senior appointments/promotions, grants panel member for NSF (USA), advisory board member for European FP6 grant centre, steering committee member for NCRIS GBROOS and SEMAT, and board member for IT and SCADA companies. His academic excellence is recognized by several invited presentations of plenary/keynote lectures and serving as General Chair, he organised over 10 international conferences.. He has served as an associate editor for journals/transactions including IEEE Transactions on Neural Networks, International Journal Distributed sensor Networks, International Journal of Biomedical Engineering. He was given a Foreign Specialist Award by the Ministry of Education, Japan. In recognition of his community and industry engagement, he received several Uni of Melb Knowledge Transfer Excellence Awards. His research interests include SVMs, sensors and sensor networks, machine learning, neural network, pattern recognition, signal processing, and control. The interdisciplinary themes for his large scale projects cover Biology, Medicine, Mathematics, Architecture, and Engineering. He is a fellow of the IEEE.

Pramod K. Varshney (S'72-M'77-SM'82-F'97) was born in Allahabad, India, on July 1, 1952. He received the B.S. degree in electrical engineering and computer science (with highest hons.), and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana-Champaign in 1972, 1974, and 1976 respectively. From 1972 to 1976, he held teaching and research assistantships at the University of Illinois. Since 1976, he has been with Syracuse University, Syracuse, NY, where he is currently a Distinguished Professor of Electrical Engineering and Computer Science and the Director of CASE: Center for Advanced Systems and Engineering. He served as the Associate Chair of the department from 1993 to 1996. He is also an Adjunct Professor of Radiology at Upstate Medical University, Syracuse, NY. His current research interests are in distributed sensor networks and data fusion, detection and estimation theory, wireless communications, image processing, radar signal processing, and remote sensing. He has published extensively. He is the author of *Distributed Detection and Data Fusion* (Springer-Verlag, 1997). He has served as a consultant to several major companies. Dr.Varshney was a James Scholar, a Bronze Tablet Senior, and a Fellow while at the University of Illinois. He is a member of Tau Beta Pi and is the recipient of the 1981 ASEE Dow Outstanding Young Faculty Award. He was elected to the grade of Fellow of the IEEE in 1997 for his contributions in the area of distributed detection and data fusion. He was the Guest Editor of the Special Issue on Data Fusion of the PROCEEDINGS OF THE IEEE, January 1997. In 2000, he received the Third Millennium Medal from the IEEE and Chancellors Citation for exceptional academic achievement at Syracuse University. He is the recipient of the IEEE 2012 Judith A. Resnik Award. He serves as a Distinguished Lecturer for the IEEE Aerospace and Electronic Systems (AES) Society. He is on the Editorial Board of the *Journal on Advances in Information Fusion*. He was the President of International Society of Information Fusion during 2001.